COMP 576 - Fall 2017
Assignment 1

September 29, 2017

Backpropagation in a Simple Neural Network

1a) Dataset

1b) Derivatives of Activation Functions

Sigmoid:

$$f(x) = \frac{1}{1+e^{-x}} = (1+e^{-x})^{-1}$$

$$\frac{d(1+e^{-x})^{-1}}{dx} = (1+e^{-x})^{-2}(e^{-x})$$

$$= \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{1+e^{-x}}\frac{e^{-x}}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\frac{(1+e^{-x})-1}{1+e^{-x}}$$

$$= \frac{1}{1+e^{-x}}\left(1 - \frac{1}{1+e^{-x}}\right)$$

$$\frac{d(1+e^{-x})^{-1}}{dx} = f(x)(1-f(x))$$

Tanh:

$$f(x) = tanh(x) = \frac{sinh(x)}{cosh(x)}$$

$$\frac{d\left(\frac{sinh(x)}{cosh(x)}\right)}{dx} = \frac{cosh(x)cosh(x) - sinh(x)sinh(x)}{cosh^2(x)}$$

$$= \frac{cosh^2(x) - sinh^2(x)}{cosh^2(x)} = \frac{cosh^2(x)}{cosh^2(x)} - \frac{sinh^2(x)}{cosh^2(x)}$$

$$\frac{d\left(\frac{sinh(x)}{cosh(x)}\right)}{dx} = 1 - tanh^2(x)$$

ReLu:

$$f(x) = max(0, x)$$

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

$$f'(x) = \begin{cases} 1, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

1c) Building the Neural Network

Three Layer Network

$$z^1 = W^1 x + b^1 \tag{1}$$
$$a^1 = actFun(z^1) \tag{2}$$
$$z^2 = W^2 a^1 + b^2 \tag{3}$$
$$a^2 = \hat{y} = softmax(z^2) \tag{4}$$

Cross Entropy of Batch

$$L(y, \hat{y}) = -\frac{1}{N} \sum_{n=1}^{N} \sum_{i=1}^{C} y_{n,i} log \hat{y}_{ni} \tag{5}$$

1d) Backward Pass - Backpropagation

Gradients